

# What Can We Learn about Country Performance from Conditional Comparisons across Countries?

*Martin Ravallion*

Existing methods for assessing latent country or institutional performance can yield deceptive results.



## Summary findings

There have been many attempts to infer latent performance attributes of governments (or other institutions) from conditional comparisons that control for observed variables. Success in doing so could greatly improve government performance.

Ravallion critically reviews the econometric foundations of the methods used. He argues that latent

heterogeneity remains a fundamental but unresolved problem.

Locating a benchmark for measuring performance adds a further problem. Current methods do not yield a consistent estimate of even the mean latent performance attribute. An assessment of country performance by these methods could well be wildly wrong.

---

This paper — a product of Poverty and Human Resources, Development Research Group — is part of a larger effort in the group to assess and improve methods for monitoring and assessing country performance. The study was funded by the Bank's Research Support Budget under the research project "Policies for Poor Areas" (RPO 681-39). Copies of this paper are available free from the World Bank, 1818 H Street NW, Washington, DC 20433. Please contact Patricia Sader, room MC4-773, telephone 202-473-3902, fax 202-522-1153, email address [psader@worldbank.org](mailto:psader@worldbank.org). Policy Research Working Papers are also posted on the Web at [www.worldbank.org/research/workingpapers](http://www.worldbank.org/research/workingpapers). The author may be contacted at [mravallion@worldbank.org](mailto:mravallion@worldbank.org). May 2000. (23 pages)

*The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent.*

# What can we learn about country performance from conditional comparisons across countries?

Martin Ravallion\*

*World Bank and Université des Sciences Sociales, Toulouse, France*

Keywords: Government performance; human development; poverty; social spending.

JEL: D61, H50, I38

---

\* Address for correspondence: [mravallion@worldbank.org](mailto:mravallion@worldbank.org) or (until 7/2000) ARQADE, Manufacture, 21 Allée de Brienne, 31 000 Toulouse. The author is grateful to Emanuela Galasso, Jeff Hammer, Jyotsna Jalan, and Dominique van de Walle for helpful comments.

performance. Section 5 then considers the large sample properties of the method when performance is measured relative to a frontier. Section 6 concludes.

## **2. Assessing performance with incomplete data**

A common practice in development policy discussions is to compare an aggregate welfare metric for a country with the metric's mean conditional on national income per capita. In an early and influential example, Sen (1981) looked at the deviations of actual social indicators from their expected values given average income. These residuals suggested that Sri Lanka was the best performer. Using a regression of (log) life expectancy on (log) income per capita, Sen showed that the expected national income corresponding to Sri Lanka's (high) life expectancy given income was about 20 times higher than the country's actual income. This excellent conditional performance was interpreted as support for Sri Lanka's high level of social spending over many decades.

Conditional comparisons of this sort have since become common in both academic literature and policy discussions. They have taken the form of either the "horizontal" comparison made by Sen (in which the difference in performance is measured in the units of the horizontal axis) or the straight "vertical" residuals. One can find many examples in studies of country performance in education and health care (see, for example, World Bank, 1993; Kakwani, 1993; Wang et al., 1999). The UNDP's "Human Development Reports" have often contained statements about country performance conditional on average income (see, for example, UNDP, 1996).

In a recent example, the WHO's "World Health Report" for 1999 provides health "performance measures" over time by country, based on the residuals obtained by regressing health aggregates on the log of GDP per capita, its squared value and a trend

(WHO, 1999, Annex Table 6). The residuals are taken to measure public performance, notably through the expansion and dissemination of knowledge about health care. It is claimed that some countries have performed considerably better than others, as assessed by this method.

The same basic idea can be used to assess the cost-effectiveness of public spending on social services in reducing poverty and inequality, recognizing that there are unobserved inefficiencies in social spending. In a recent example, Gouyette and Pestieau (1999) compare measures of poverty and inequality with levels of social spending across OECD countries, and they use this comparison to form judgements about the efficiency of the welfare state.

The application found in Gouyette and Pestieau provides a potential model for further work of this sort, extended to include developing countries. This offers hope for measuring latent country effort in fighting poverty, thereby enhancing (dramatically, it would seem) the scope for using external development assistance to encourage country action to reduce poverty. Ostensibly, this method turns a potentially difficult agency problem (with hidden information on government effort) into a relatively simple contract. For this reason, and also to help make the discussion more concrete, I will draw on the Gouyette and Pestieau application repeatedly in this paper. However, this is not a comment on Gouyette and Pestieau; the generic method is more common than this one paper (and that paper contains much of interest that I do not touch on here).

Indeed, the method has not been confined to country comparisons. The same basic idea has been used to assess the performance of business schools (Hamlen and Southwick, 1989; Tracy and Waldfogel, 1997). Here the performance indicator is the

average starting salary of graduates, and the controls are measures of student quality (test scores and experience). The residuals are taken to measure the school's "value added".

Naturally there have been some disputes in the literature about how the precise way such conditional comparisons should be made. In the development literature, there have been concerns about the functional form of the measures used (as in Kakwani's, 1993, analysis of how to measure achievements in health development). And there has been an issue about whether one should use levels or growth rates (as in Bhalla and Glewwe, 1986; Aturupane et al., 1994; Ranis et al., 2000). But all these methods share the common feature of using conditional comparisons to assess latent aspects of performance.

Some of the specific conditional comparisons have been controversial. For example, Sen's (1981) claim that Sri Lanka was an outstanding performer in social policies given its income generated considerable debate (see Bhalla and Glewwe, 1986, and Sen's, 1988, reply).<sup>1</sup> However, subsequent analysis of time series data for Sri Lanka has demonstrated a significant and quantitatively important role of social spending in reducing infant mortality at given average income (Anand and Ravallion, 1993).

And there are often concerns about the narrowness of the welfare objective assumed by the method in practice. For example, reducing poverty is not the only objective of social spending (as Gouyette and Pestieau, 1999, note). Judging the efficiency of the welfare state by only one of its objectives is hazardous. And even if poverty was the sole objective, there are a number of issues begging about how this

---

<sup>1</sup> For a review of this debate see Kakwani (1993). Also see Anand and Ravallion (1993) and Aturupane et al. (1994).

should be measured in this context. However, I shall put these data problems to one side, and focus on the reliability of the conditional comparison method for an agreed objective.

The right way to interpret these conditional measures of country performance has never been entirely clear in this literature. A country that appears to be performing poorly now in terms of its average health attainments conditional on income could become a star performer in the future simply by (suffering) negative growth, without any gain in actual health attainments (Ravallion, 1997). This is a nagging concern about the country comparisons of health care performance found in WHO (1999).

However, the literature has tended to justify these methods on intuitive grounds. We still do not have a clear idea of the conditions under which they will be reliable.

### **3. A case study on measuring the efficiency of social spending**

There is a large literature on how the welfare impacts of public spending on social services can be assessed empirically with micro data.<sup>2</sup> A prominent issue has been the possible biases involved (notably in the common forms of non-experimental evaluation). The methods considered here take a far more aggregated approach than is common in the evaluation literature. But (as I will discuss later) there are similar concerns about biases in using cross-country comparisons to make statements about country performance, though the source of bias depends on specifics of each method.

In their use of the conditional comparison method in assessing the efficiency of social spending, Gouyette and Pestieau (1999) find wide differences in poverty rates conditional on social spending in a sample of OECD countries. Their measure of poverty is the percentage of the population with adjusted household income less than half the

median (following a common practice in Western Europe). They point to the example of France, which has about the same social spending as Belgium (26% of GDP), yet a poverty rate that is three times higher. If France had eliminated the waste in its social spending then (*ceteris paribus*) it could achieve a poverty rate of 3% with its current spending — instead of 9%. Two thirds of France's poverty is attributed to its inefficient social spending. If this is right, the policy implications are of great interest.

A problem with conditional comparison methods in practice has been to come up with a cardinal measure of the value of the latent performance attribute. Looking at the residuals in the regression of observed performance on its observed determinants, practitioners of this method make claims about how one country is doing relative to another. But one would also like to have a measure of absolute performance.

The types of “horizontal” comparisons made by Sen (1981) provide a simple monetary measure of performance. Gouyette and Pestieau (1999) build on this method by using their regressions to locate an “efficiency frontier” for the welfare state. They locate the frontier by regressing their poverty (or inequality) measure on social spending and then shifting the regression line down to find the country with lowest poverty conditional on social spending i.e., the largest (negative) residual. They then interpret the difference between actual spending and its expected value on this frontier as the amount of “inefficiency” in social spending. I will call this the “Efficiency Frontier Method” (EFM).

Figure 1 illustrates the method. The ellipse encompasses the scatter points of the data, and the upper straight line is the regression line; the lower one is the estimated frontier, passing through the data point for country *B*, with the largest residual. Mean

---

<sup>2</sup> A useful compilation of papers following various approaches, with critical commentary by the editors, can be found in van de Walle and Nead (1995).



social spending is  $\bar{S}$ , while the estimated mean level of efficient social spending is  $\hat{S}_E$  corresponding to the average level of poverty ( $\bar{P}$ ) on the frontier. The estimated mean level of inefficient spending is the difference between  $\bar{S}$  and  $\hat{S}_E$ .

Across their OECD sample, Gouyette and Pestieau find a (highly) significant negative correlation between social spending and poverty, as measured by the proportion of people with household adjusted income below half the median. They also find a significant negative correlation with the Gini index of income inequality. On constructing their efficiency frontiers from these regressions, they present graphs of poverty and inequality in OECD countries against social spending. Belgium turns out to be the best performer in their main analysis. The results indicate some large inefficiencies, such that much greater welfare impacts could be achieved at the same budgetary cost, or that the cost of the same impact could be greatly reduced.

Gouyette and Pestieau do not give the implied levels of inefficient social spending by country. But it is not difficult to make the calculations. Table 1 gives the share of social spending that is deemed “inefficient”.<sup>3</sup> The average level of inefficient social spending is one third of total social spending, or about 9% of GDP. There is considerable variation. For Holland, about half of social spending is deemed inefficient, while for Belgium it is zero (by construction) and for Sweden it is under 20%.

All this is a natural extension of long-standing methods found in the literature whereby cross-country comparisons conditional on the mean of one or more relevant control variables (average incomes or public spending, for example) are used to make

---

<sup>3</sup> The calculations correspond to Figure 2 in Gouyette and Pestieau. I did not get the same coefficients they report, but the difference is small. My estimate of the regression coefficient of log poverty on log social spending was  $-1.71$  ( $t=-4.23$ ) with an intercept of  $7.49$  ( $t=5.71$ ).

statements about country performance. Belgium plays a similar role in the analysis of Gouyette and Pestieau (1999) as Sri Lanka played in Sen (1981). The following sections examine the properties of these methods, and their reliability for guiding policy.

#### 4. Using conditional comparisons to assess country performance

The types of cross-country performance comparisons discussed in the last two sections can be applied to many types of policies, and they can yield seemingly strong results about otherwise unobserved aspects of country performance. The fact that these are such simple and seemingly powerful methods, that influence policy discussions, makes it compelling to take a closer look at their foundations.

Continuing with the example of the problem addressed by Gouyette and Pestieau, “inefficient social spending” in this context can be defined as social spending that has no (direct or indirect) impact on poverty or inequality. Observed social spending combines two components — one that reduces poverty and one that does not. The underlying true regression gives poverty (or inequality) as a decreasing function of the first component, but (by definition) the second has no effect. However, the estimated regression cannot make this distinction, and so it is total social spending that appears on the right hand side.

Let the two components of total social spending ( $S$ ) be  $S^E$ , the efficient component, and  $S^I$ , the inefficient component; the level of poverty depends on  $S^E$ , but not  $S^I$ . Let  $\alpha + \beta S_i^E$  give the expected level of poverty in country  $i$  given efficient social spending  $S_i^E$ , i.e., the actual level of poverty is:

$$P_i = \alpha + \beta S_i^E + \varepsilon_i \quad (1)$$

where  $\alpha$  and  $\beta$  are parameters and  $\varepsilon_i$  is a zero-mean i.i.d. error term. We do not observe the amount of efficient social spending but only total social spending:

$$S_i = S_i^E + S_i^I \quad (2)$$

Consider France again. The actual level of poverty in France is:

$$P_F = \alpha + \beta S_F^E + \varepsilon_F \quad (3)$$

Belgium, on the other hand has the same level of social spending but a much lower poverty rate. The relevant counter-factual level of poverty in France is found by replacing its wasteful social spending with the efficient social spending found in Belgium, while holding all else constant. So efficient social spending in France rises from  $S_F^E$  to  $S_F = S_B$ . This new level of poverty in France is given by:

$$P_F^* = \alpha + \beta S_B + \varepsilon_F \quad (4)$$

The amount by which poverty in France would fall is then:

$$P_F - P_F^* \equiv (P_F - P_B) - (\varepsilon_F - \varepsilon_B) \quad (5)$$

(Noting that  $P_B = \alpha + \beta S_B + \varepsilon_B$  since Belgium is deemed to be on the efficiency frontier.) So to calculate the extra poverty in France due to inefficient social spending we have to net out the difference in poverty rates conditional on efficient social spending.

To be correct, the inference drawn from this conditional country comparison requires that there are no differences in poverty between countries once one controls for differences in levels of efficient social spending not total social spending. However, one can think of many reasons for differences in poverty and inequality between countries besides differences in efficient social spending. Differences in the progressivity of the taxes used to finance social spending, and in labor market policies are examples.

Notice that (so far) the regression model takes a back seat role. We did not of course need the regression to compare France and Belgium, since they happen to have a very similar level of social spending. Other applications of this type of conditional comparison use the regression to establish the conditional mean, at any country's actual level of spending. This is then used as the benchmark for judging performance. Bias in the regression parameters is not of concern; all we want is a good estimate of the conditional mean. One need not even use a parametric model for this purpose.

However, the key point remains that controlling for the observed level of spending does not allow one to identify the latent difference in the efficiency of spending. Other latent variables and measurement errors create heterogeneity; poverty rates naturally vary at given levels of efficient social spending. We can call this the "heterogeneity bias" in conditional comparisons of country performance.

Heterogeneity bias will become less severe the more controls for observable heterogeneity that one adds. Instead of comparing poverty in France with Belgium (with the only control being for their observed social spending) suppose we compare it with the expected poverty level in Belgium if it had the other observed characteristics of France. As long as those characteristics have some explanatory power when added to the regression of poverty on social spending, then this modified conditional comparison will be more revealing of the latent difference in the efficiency of social spending.

For example, the poverty rate in OECD countries may well also depend on the scale of active labor market programs. Such programs in Belgium accounted for 1.21% of GDP in 1985-91, versus 0.77% in France (Vanhoudt, 1997, quoting OECD data). On adding the share of GDP devoted to active labor market programs to the regression of

poverty on social spending I find that both variables have poverty reducing effects and are significant. The regression coefficient on social spending is  $-0.37$  and that on active labor market programs is  $-2.90$  (with t-ratios based on White standard errors of 2.85 and 2.12 respectively). (Using a double log specification the coefficients on log social spending and active labor market programs are  $-1.04$  and  $-0.50$ , with t-ratios of 2.50 and 2.11.) Since Belgium devoted a larger share of GDP to these programs than did France, this accounts in part for the difference in poverty; about 1.3 percentage points of the difference can be explained this way. No doubt other observable sources of heterogeneity could be found to help bridge the gap.

This type of latent heterogeneity is familiar from the literature on program evaluation. Controlling for observed performance attributes does not assure unbiased estimates of program impact, though there is evidence that with careful choice of controls (such as using propensity score matching methods) the bias can be greatly reduced (Heckman et al., 1998). The low dimensionality of the controls used in past conditional country comparisons does not, however, leave one with much confidence.

## **5. Large sample properties**

Sizable errors in estimating a latent performance attribute in a specific country cannot be ruled out. How do these methods perform in the aggregate for large samples?

When the benchmark is the conditional mean, and this is estimated consistently, then the heterogeneity bias in a comparison of any one country's performance with that of the benchmark country will persist no matter how large the sample. The average deviation from this benchmark will obviously converge in probability to zero as the sample size increases. That merely tells us how uninteresting that benchmark is in

measuring aggregate performance, which helps motivate interest locating an efficiency frontier. So this section will focus on the large sample properties of the EFM, continuing to use as an example the specific application found in Gouyette and Pestieau (1999).

Consider again the regression specification of poverty on social spending in equation (1); this can be written as:

$$P_i = \alpha + \beta S_i + \mu_i \quad (i=1, \dots, n) \quad (6.1)$$

$$\mu_i = -\beta S_i^I + \varepsilon_i \quad (6.2)$$

This has a superficial resemblance to the stochastic frontier production model in which the error term in an empirical production or cost function combines a one-sided disturbance term (representing inefficiency) as well as a regular zero mean error (Aigner et al., 1977). However, there is an important difference, namely that in the present case the one-sided component is correlated with the regressors. Existing methods for estimating stochastic frontier models on cross-sectional data do not apply to this case.

In virtually all of the conditional comparison methods found in the literature, equation (6.1) is estimated using ordinary least squares. (I discuss the problems in doing so later.) The EFM then estimates the level of efficient social spending in country  $i$  by the value of its spending if it were on the “efficiency frontier” obtained by shifting the intercept of their regression until it passes through the data point for the country with the largest (negative) residual ( $\hat{\mu}_B$ ). In other words, the estimate of efficient social spending for country  $i$  ( $\hat{S}_i^E$ ) is obtained by inverting the following equation:

$$P_i = \hat{\alpha} + \hat{\beta} \hat{S}_i^E + \hat{\mu}_B \quad (7)$$

where  $\hat{\alpha}$  and  $\hat{\beta}$  are their least squares estimates of the parameters of (6.1). It can be seen from (7) that the estimate of efficient social spending in any country is nothing more than a fixed linear transform of its observed poverty measure. By definition:

$$P_i = \hat{\alpha} + \hat{\beta} S_i + \hat{\mu}_i \quad (8)$$

Subtracting (7) from (8), the estimated level of inefficient spending is:

$$\hat{S}_i^I = S_i - \hat{S}_i^E = \frac{\hat{\mu}_B - \hat{\mu}_i}{\hat{\beta}} = s_i - s_B + \frac{(p_B - p_i)}{\hat{\beta}} \quad (9)$$

where the lower case letters denote deviations from sample means. As  $n$  goes to infinity, the sample mean of  $\hat{S}_i^I$  will converge in probability to:

$$\text{plim}(\sum \hat{S}_i^I / n) = \frac{\hat{\mu}_B}{\text{plim}(\hat{\beta})} = -s_B + \frac{p_B}{\text{plim}(\hat{\beta})} \quad (10)$$

How does this compare to the true mean of inefficient social spending? Taking expectations through (6.2),  $E(\mu) = -\beta E(S^I)$ ; this is the expected value of the decrease in poverty if inefficient social spending was replaced by efficient social spending. So:

$$E(S^I) = \frac{-E(\mu)}{\beta} = -s_B + \frac{p_B - \mu_B}{\beta} \quad (11)$$

Comparing (10) and (11), the asymptotic bias in the estimate of average level of inefficient social spending is then given by:

$$\text{Bias} = p_B \left[ \frac{1}{\text{plim}(\hat{\beta})} - \frac{1}{\beta} \right] + \frac{\mu_B}{\beta} \quad (12)$$

There are two sources of bias. The first relates to the bias in the regression coefficient, leading to miss-identification of the shape of the efficiency frontier. The second relates to

unobserved variables that influence poverty in the benchmark country at given social spending i.e., miss-identification of the height of the frontier.

Consider the first source of bias. If one accepts that there is inefficiency in social spending ( $E(S^I) > 0$ ) on a priori grounds, then the OLS estimate of  $\beta$  is automatically inconsistent. If wasteful social spending is uncorrelated with efficient social spending then this is nothing more than a classic attenuation bias due to measurement error in a regressor. The OLS regression coefficient of  $P$  on  $S$  converges in probability to  $\beta \text{var}(S^E) / \text{var}(S)$ . Then (12) simplifies to:

$$\text{Bias} = \frac{p_B \text{var}(S^I)}{\beta \text{var}(S^E)} + \frac{\mu_B}{\beta} \quad (13)$$

The first term on the RHS must be positive if social spending reduces poverty ( $\beta < 0$ ) and the benchmark country has lower poverty than average ( $p_B < 0$ ).

With respect to the second source of bias, notice that assuming that social spending in the benchmark country is entirely efficient ( $S_B^I = 0$ ) still leaves the bias due to other unobserved variables influencing poverty in the benchmark country ( $\varepsilon_B \neq 0$ ). There is no obvious a priori argument one could make about this source of bias, and so the overall direction of bias is theoretically ambiguous. If the benchmark country has favorable conditions for poverty reduction, at given social spending ( $\varepsilon_B < 0$ ) then this second source of bias will reinforce the bias arising from the regression miss-specification when the benchmark country has below average poverty. For example, as we saw in Section 3, Belgium's active labor market policies tend to be poverty reducing at given social spending; this effect would add to the bias in the estimate of aggregate



inefficiency by leading the method to miss-locate the efficiency frontier. However, if the benchmark country has sufficiently unfavorable conditions then the method will underestimate the extent of inefficiency in social spending.

Figure 2 illustrates the case when the benchmark country has favorable conditions for reducing poverty. The bold line is the true efficiency frontier, with a steeper slope (the first source of bias) and at a higher level (the second source) than the estimated frontier. The way the figure is drawn, the EFM's estimate of mean inefficient social spending is roughly double the true mean (i.e., the distance from X to Z is roughly double that from Y to Z).

To make the discussion concrete, the above analysis has focused on the application in Gouyette and Pestieau (1999). This is only one of many possibilities. In keeping with the literature on human development (discussed in section 2), one could reinterpret the above analysis, reading " $P$ " as a social indicator such as infant mortality, and " $S$ " as GDP per capita, some of which is devoted to things that are good for child health ( $S^E$ ) while some is not ( $S^I$ ). For example, it has been argued that aggregate health indicators such as the infant mortality rate depend far more on incomes of the poor than the nonpoor (Bidani and Ravallion, 1997), so country performance in income poverty reduction matters to human development. (Indeed, along with the provision of key social services, this may well account fully for the relationship with average income; see Anand and Ravallion, 1993). Some country (possibly Sri Lanka) will anchor the frontier, and one could then measure the distance from this frontier to the regression line, or any country's data point, as a measure of performance relative to this welfare

objective. How reliable a guide to country performance all this will be will then depend on the factors I have identified above.

## **6. Conclusions**

Empirical economists normally ignore the residuals, and focus on the regression line (though often using the residuals in testing the regression). A strand of empirical work has tried to learn something more from the residuals. The starting point is the observation that cross-country differences in aggregate welfare measures conditional on observed welfare determinants will contain the effects of latent determinants of interest, notably the unobserved efforts of governments. But they will contain a lot more than that, including measurement errors.

Merely controlling for the observed determinants does not, in general, allow one to infer the effect of just one of the unobservables, or any subset. This is an obvious but fundamental identification problem in all such methods of using cross-country comparisons to assess specific latent aspects of government performance. Nor can one rely on large samples to get rid of the problem.

Measuring performance relative to an “efficiency frontier” adds a further source of bias, via the model used to locate the frontier. Total social spending (for example) must surely be correlated with its own components, including both the efficient and inefficient types of spending. So the error term in a regression of poverty on social spending is very likely to be correlated with the regressor, rendering least squares estimates (including the residuals) biased and inconsistent. This leads to miss-identification of the shape of the frontier, which pushes the method toward overestimation of the inefficiency of social spending. Similarly, if one is using this

method to assess country performance in child health policies by looking at the residuals from a regression of the infant mortality rate on average income, there is likely to be a bias due to correlation between the latent policy variables and income. Unlike the standard conditional comparisons, bias in the regression parameters now introduces a bias in the inferences drawn about country performance, because the analyst gets the slope of the efficiency frontier systematically wrong.

Latent heterogeneity is also a problem in locating the efficiency frontier. There will be latent determinants of income distribution in the best-case “frontier” countries that have nothing to do with social spending policies. Favorable (unfavorable) latent conditions in the frontier countries will lead the frontier method to overestimate (underestimate) the potential for improving policies.

The direction of bias in estimates of latent policy variables cannot be determined on theoretical grounds. Indeed, it is not clear what can be inferred about the aggregate cost-effectiveness of public policies even in large samples. And it is dangerous indeed to try to use this type of method to say something about the performance of a specific country in a sample of any size.

Adding extra control variables will help, though even with lots of controls, one must take seriously the problem that the unobserved and observed attributes of performance are unlikely to be orthogonal. Any independent evidence on the reliability of observed performance indicators would also help by providing information about the likely biases in least squares estimates. But it is not immediately obvious where this evidence would come from in practice. Alternatively, one might look for valid instrumental variables, though it is not clear what would qualify in this context. Panel

data would not help much in the settings in which this problem arises, since assuming that the latent performance attribute is time invariant (and identified) would greatly diminishing the method's attraction as a tool for monitoring performance over time.

So this review must end on a rather negative note. There appears to be little hope of using the existing methods found in the literature to calibrate a reliable measure of latent country or institutional performance.

## References

- Aigner, D., C. Lovell and P. Schmidt (1977), Formation and estimation of stochastic frontier production models, *Journal of Econometrics* Vol. 6, pp. 21-37.
- Anand, Sudhir and Martin Ravallion (1993), Human development in poor countries: On the role of private incomes and public services, *Journal of Economic Perspectives* Vol. 7, pp.133-150.
- Aturupane, Harsha, Paul Glewwe and Paul Isenman (1994), Poverty, human development and growth: An emerging consensus? *American Economic Review, Papers and Proceedings* Vol. 84(2), pp. 244-249.
- Bhalla, Surjit S., and Paul Glewwe (1986), Growth and equity in developing countries: A reinterpretation of the Sri Lankan experience, *World Bank Economic Review* Vol. 1, pp. 35-63.
- Bidani, Benu and Martin Ravallion (1997), Decomposing social indicators using distributional data, *Journal of Econometrics* Vol. 77(1), pp. 125-140.
- Gouyette, Claudine and Pierre Pestieau (1999), Efficiency of the welfare state, *Kyklos*. Vol. 52, pp. 537-553.
- Hamlen, W. and L. Southwick (1989), Quality in the MBA program: Inputs, outputs or value added? *Journal of Economic and Social Measurement* Vol. 15(1), pp.1-26.
- Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998), Characterizing selection bias using experimental data, *Econometrica* Vol. 66, pp.1017-1099.
- Jamison, D.T., J.Wang, K. Hill and J.L. Londono (1996), Income, mortality and fertility in Latin America: Country level performance, 1960-90, *Revista-de-Analisis-Economico*. Vol. 11(2), pp. 219-61.
- Kakwani, Nanak (1993), Performance in living standards: An international comparison, *Journal of Development Economics* Vol. 41, pp. 307-336.
- Ranis, Gustav, Frances Stewart and Alejandro Ramirez (2000), Economic growth and human development, *World Development* Vol. 28(2), pp. 197-220.
- Ravallion, Martin (1997), Good and bad growth: The Human Development Reports, *World Development* Vol. 25(5), pp. 631-638.
- Sen, Amartya K., (1981), Public action and the quality of life in developing countries, *Oxford Bulletin of Economics and Statistics*, Vol. 43, pp. 287-319.

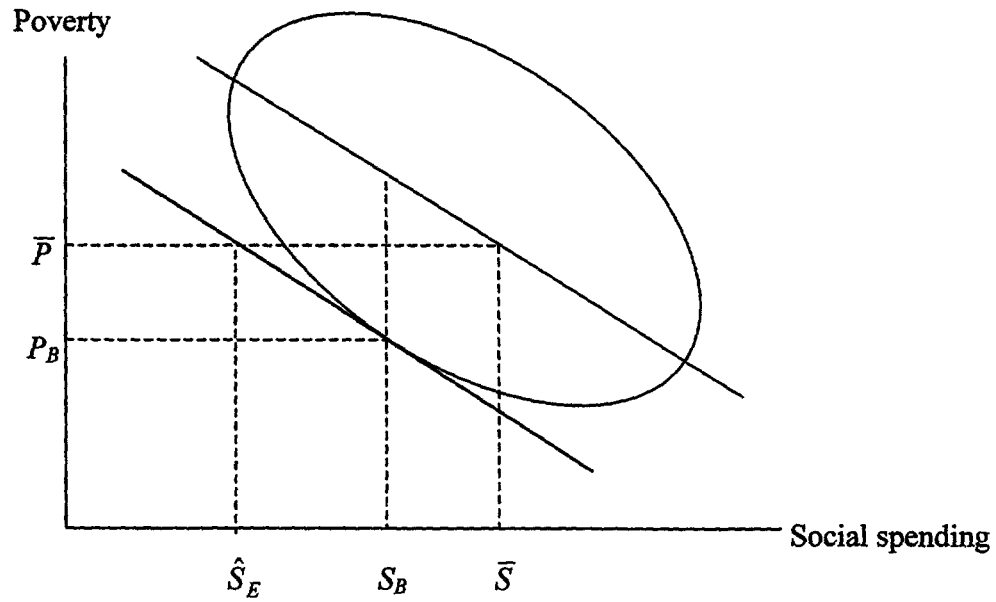
- \_\_\_\_\_, (1988), Sri Lanka's achievements: When and how?, in Srinivasan, T.N., and P.K. Bardhan (eds) *Rural Poverty in South Asia*, New York: Columbia University Press, pp. 549-56.
- Tracy, J. and J. Waldfogel, (1997), The best business schools: A market based approach, *Journal of Business*, Vol. 70, pp.1-31.
- United Nations Development Programme (UNDP) (1996), *Human Development Report*. New York: Oxford University Press.
- van de Walle, Dominique and Kimberly Nead (eds) (1995), *Public Expenditures and the Poor*, Baltimore: Johns Hopkins University Press.
- Vanhoudt, Patrick (1997), Do labor market policies and growth fundamentals matter for income inequality in OECD countries?, *IMF Staff Papers*, Vol. 44(3), pp. 356-373.
- Wang, Jia, Dean T. Jamison, Eduard Bos, Alexander Preker, John Peabody (1999), *Measuring Country Performance on Health: Selected Indicators for 115 Countries*, Health, Nutrition and Population Series, World Bank.
- World Bank (1993), *World Development Report: Investing in Health*, New York: Oxford University Press.
- World Health Organization (1999), *The World Health Report: Making a Difference*. Geneva: World Health Organization.

**Table 1: Total social spending and estimated inefficient social spending**

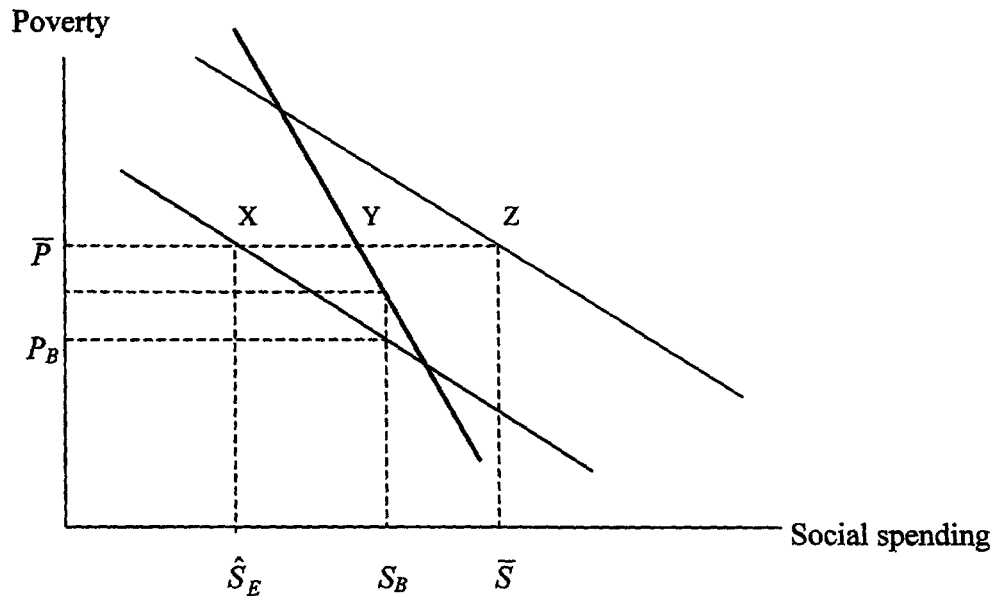
	Poverty rate (% below half median adjusted income)	Social spending (% of GDP) 1991-93 mean	Inefficient social spending, estimated by the Efficiency Frontier Method (% of GDP)
Belgium	3.3	25.7	0.0
Denmark	6.3	31.1	13.5
France	8.7	26.2	11.6
Germany	6.2	26.7	8.9
Ireland	8.1	21.4	6.2
Italy	9.8	23.3	9.7
Netherlands	7.2	30.9	14.6
Spain	9.1	19.7	5.5
United Kingdom	10.8	24.3	11.4
Austria	8.6	24.0	9.3
Finland	4.6	30.6	9.4
Sweden	2.5	37.0	6.8
United States	17.2	14.8	5.0
Mean	7.9	25.8	8.6

Note: Data on poverty and social spending are from Gouyette and Pestieau (1999) who cite the Luxembourg Income Study and the OECD as the sources.

*Figure 1*  
The Efficiency Frontier Method of Assessing Social Spending



*Figure 2*  
Bias in the Efficiency Frontier Method





# Policy Research Working Paper Series

	<b>Title</b>	<b>Author</b>	<b>Date</b>	<b>Contact for paper</b>
WPS2328	Social Transfers and Social Assistance: An Empirical Analysis Using Latvian Household Survey Data	Branko Milanovic	April 2000	P. Sader 33902
WPS2329	Improving Russia's Policy on Foreign Direct Investment	Joel Bergsman Harry G. Broadman Vladimir Drebensov	May 2000	S. Craig 33160
WPS2330	Reducing Structural Dominance and Entry Barriers in Russian Industry	Harry G. Broadman	May 2000	S. Craig 33160
WPS2331	Competition, Corporate Governance, and Regulation in Central Asia: Uzbekistan's Structural Reform Challenges	Harry G. Broadman	May 2000	S. Craig 33160
WPS2332	Financial Intermediary Distress in the Republic of Korea: Small Is Beautiful?	Paola Bongini Giovanni Ferri Tae Soo Kang	May 2000	K. Labrie 31001
WPS2333	Output Fluctuations in Latin America: What Explains the Recent Slowdown?	Santiago Herrera Guillermo Perry Neile Quintero	May 2000	C. Palarca 35328
WPS2334	Sex Workers and the Cost of Safe Sex: The Compensating Differential for Condom Use in Calcutta	Vijayendra Rao Indrani Gupta Smarajit Jana	May 2000	P. Sader 33902
WPS2335	Inflation and the Poor	William Easterly Stanley Fischer	May 2000	K. Labrie 31001
WPS2336	Endogenous Enforcement and Effectiveness of China's Pollution Levy System	Hua Wang David Wheeler	May 2000	Y. D'Souza 31449
WPS2337	Pollution Charges, Community Pressure, and Abatement Cost of Industrial Pollution in China	Hua Wang	May 2000	Y. D'Souza 31449
WPS2338	The Geography of International Investment	Howard J. Shatz Anthony J. Venables	May 2000	L. Tabada 36896
WPS2339	Building Subnational Debt Markets in Developing and Transition Economies: A Framework for Analysis, Policy Reform, and Assistance Strategy	Michel Noel	May 2000	M. Noel 32581

# Policy Research Working Paper Series

Title	Author	Date	Contact for paper
WPS2340 Currency Substitution in Latin America: Lessons from the 1990s	Pere Gomis-Porqueras Carlos Serrano Alejandro Somuano	May 2000	M. Puentes 39621
WPS2341 The Tyranny of Concepts: CUDIE (Cumulated, Depreciated Investment Effort) Is <i>Not</i> Capital	Lant Pritchett	May 2000	R. Widuri